DOCUMENT RESUME

ED 426 451 EA 029 474

AUTHOR	Dorn, Sherman
TITLE	The Political Legacy of School Accountability Systems.
INSTITUTION	Arizona State Univ., Tempe. Coll. of Education.
ISSN	ISSN-1068-2341
PUB DATE	1998-01-00
NOTE	32p.; Electronic journal article. Made available via the EPAA web site (cited below).
AVAILABLE FROM	Web site: http:/olam.ed.asu.edu/epaa/v6n1.html
PUB TYPE	Guides - Non-Classroom (055) -- Journal Articles (080)
JOURNAL CIT	Education Policy Analysis Archives; v6 n1 1998
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Standards; *Accountability; Educational Change; Elementary Secondary Education; Outcomes of Education; *Politics of Education; School Policy; *Standardized Tests

ABSTRACT

Recent battles over proposed national testing programs do not tell the important political story about high-stakes tests. An overview of the politics of school-accountability systems is offered in this World Wide Web journal article. Politically popular, school-accountability systems in many states already revolve around statistical results of testing with high-stakes environments, meaning that the future of high-stakes tests does not depend on what happens in Washington. Rather, the existence of tests depends largely on the political culture of published test results. Most critics of high-stakes testing do not talk about that culture. They typically focus on the "practice legacy" of testing, and the ways in which testing creates perverse incentives against good teaching. More important may be the "political legacy," or how testing defines legitimate discussion about school politics. The consequence of statistical accountability systems will be the narrowing of purpose for schools, impatience with reform, and the continuing erosion of political support for publicly funded schools. Dissent from the high-stakes accountability regime that has developed around standardized testing, including proposals for professionalism and performance assessment, commonly fails to consider these political legacies. Alternatives to standardized testing that do not also connect schooling with the public at large will be politically unviable. (Author/RJM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 426 451

EPAA • enter the archives • browse the abstracts • the editors • the edit board
• submit article • submit commentary • search • subscribe
volume: • 1 • 2 • 3 • 4 • 5 • 6 • 7

This article has been retrieved 3719 times since January 2, 1998.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Education Policy Analysis Archives

Volume 6 Number 1 January 2, 1998 ISSN 1068-2341

A peer-reviewed scholarly electronic journal. Editor: Gene V Glass Glass@ASU.EDU. College of Education Arizona State University,Tempe AZ 85287-2411 Copyright 1998, the EDUCATION POLICY ANALYSIS ARCHIVES.Permission is hereby granted to copy any article provided that EDUCATION POLICY ANALYSIS ARCHIVES is credited and copies are not sold.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
G. Glass
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

The Political Legacy of School Accountability Systems

Sherman Dorn
University of South Florida

Abstract

The recent battle reported from Washington about proposed national testing program does not tell the most important political story about high stakes tests. Politically popular school accountability systems in many states already revolve around statistical results of testing with high-stakes environments. The future of high stakes tests thus does not depend on what happens on Capitol Hill. Rather, the existence of tests depends largely on the political culture of published test results. Most critics of high-stakes testing do not talk about that culture, however. They typically focus on the *practice legacy* of testing, the ways in which testing creates perverse incentives against good teaching. More important may be the *political legacy* , or how testing defines legitimate discussion about school politics. The consequence of statistical accountability systems will be the narrowing of purpose for schools, impatience with reform, and the continuing erosion of political support for publicly funded schools. Dissent from the high-stakes accountability regime that has developed around standardized testing, including proposals for professionalism and performance assessment, commonly fails to consider these political legacies. Alternatives to standardized testing which do not also connect schooling with the public at large will not be politically viable.

Introduction

The short-term question about high-stakes testing is not whether it shall prevail but who shall control it. The president of the United States advocates the use of standardized testing developed by the federal government. (Note 1. Opens in separate browser window.) Conservatives who vigorously oppose nationalized curriculum and testing agree that testing should exist, but organized on a state and local level instead (see Diegmueller and Lawton 1996; Lawton 1997). The recent compromise between Rep. William Goodling and the White House left the long-term fate of a truly national testing program unresolved (Hoff

EA 029 174

1997). Nonetheless, what is not at stake is the existence of high-stakes testing. Recent polling suggests that the idea of national testing is very popular (Rose, Gallup and Elam 1997), and that popularity reflects the past twenty years' growth of standardized testing. The debate over the control of testing takes for granted the existence of standardized testing because of its recent history. States for many years have been accumulating testing requirements which their legislatures, state officials, or local administrators have chosen. Despite considerable evidence that high-stakes testing distorts teaching and does not give very stable information about school performance, test results have become the dominant way states, politicians, and newspapers describe the performance of schools. Some have continued to note the problems of high-stakes standardized testing (e.g., Madaus 1991; McGill-Franzen and Allington 1993; Neill 1996; Noble and Smith 1994; Shepard 1991; Smith 1991; Smith and Rottenberg 1991; Wirth 1992: Chap. 7). Others try to accommodate some measure of standardized testing while building what they see as safeguards against obvious abuses. Still others (administrators in systems or schools with above-average test scores) use results as part of a marketing or public relations strategy. Few critics of high-stakes testing, however, have explicitly noted the way in which the public use of accountability systems shapes the politics of education writ large.

Statistical accountability systems are important because numbers have visible power in public debate. Anyone who listens to or reads politicians, journalists, and social critics will hear statistical references. Slowly over the last century, statistics have taken a prominent place in political culture. Whether the statistic is the official unemployment rate, poverty rates, poll results, or SAT scores, a specific number fills a niche in discussion. As Carol Weiss (1988: 168) wrote,

> The media report the proportion of the population that has been out of work for fifteen weeks or more, characteristics of high schools which have the highest drop-out rates, reasons given by voters for choosing candidates. These kinds of data become accessible and help to inform policy debates.

A number connotes objectivity or, at the very least, legitimacy. Because we perceive numbers and statistics as having a certain force on its face (just by being quantitative), we allow statistics to shape our perception of the world and the issues we perceive as important. They present selective information and thus center discussion around specific topics (silencing others). Nonetheless, we often yearn for the end of political uncertainty through statistics. Partisans in a conflict may heatedly argue that their methods are better, or their opponents' use of statistics is politically motivated, yet behind the veneer of cynicism lurks a desire for unquestionable statistics that will end debate. Maybe the official poverty line is arbitrary, but others have calculated alternative poverty estimates (Axinn and Stern 1988: 73-77; Ruggles 1990). The portrayal of a "rising tide of mediocrity" in schools was an alleged lie, but then the critics presented their own statistics as counter-evidence (Berliner and Biddle 1995; Bracey 1991, 1992, 1993, 1994, 1995a, 1996, 1997; National Commission on Excellence in Education 1983).

The production and presentation of statistics is part of the fabric of public debate, and public policy that involves the heavy use of statistics must consider the long-term consequences of that use. At least two such consequences are important, what I will call the *practice* and *political* legacies of statistics. The distinction between the two revolves around related but heuristically distinct issues:

- How do policies based on statistics shape practice?
- How do policies based on statistics shape future public policy debate?

The *practice legacy* of statistics is the nuts and bolts of how statistics shape government and private action. For example, the official U.S. consumer price index determines cost-of-

living indices for Social Security, government pay schedules, and the behavior of many private organizations. Census population counts determine state representation in the U.S. House of Representatives and some federal spending patterns. This practice legacy can, by itself, engender vivid disagreement about statistical mechanisms. In 1997, several so-called deficit hawks suggested changing the calculation of the consumer price index to lower cost-of-living indices deliberately. While they claimed that the official inflation statistics misrepresented the "true" amount of inflation, reporters and groups such as the United Auto Workers clearly understood that the argument was not about the most accurate picture of inflation but was, in large part, about the practice legacy of inflation statistics for the U.S. federal budget, entitlement programs, and private company wages and benefits (e.g., "Will Washington Cut Our COLA?" 1997). Similarly, debate about the conduct of the decennial U.S. census in the past ten years has revolved not around accuracy but policy consequences. If, as some have proposed, the Bureau of the Census augments its population count with samples to measure undercounting and adjusts the official counts with the help of samples, the distribution of federal aid to cities and states as well as Congressional representation will change according to adjustment for undercounting. Politicians in jurisdictions with alleged undercounting have an interest in supporting such adjustment based on sampling because adjusted population counts would give their constituencies higher federal aid. Other politicians have an equally intense incentive in opposing the use of sampling to prevent the loss of federal aid (Mears 1997; Roush 1996). The practice legacy of statistics is an obvious consequence of tying statistics to public policy. The examples above show specific practice legacies, when statistics are mechanisms of what Paul Starr (1987: 55-57) calls "automatic pilots." They may be less obvious in the creation of systems of incentives, as some argue that high-stakes testing environments create. Whether the result is from explicit formulae or a consequence of incentives, a practice legacy is the influence of policy on short-term behavior.

What is less clear, but equally important, is the *political legacy* of statistics, the way that the use of statistics by itself shapes public debate. (Note 2. Uses second browser window.) Discussion about teenage pregnancy is a good example of how the existence and distribution of statistics shapes debate. In the late 1960s and early 1970s, as teenage birth rates were decreasing, the Alan Guttmacher Institute and others began publicizing estimates of teen fertility statistics to illustrate what they termed an epidemic of teenage pregnancy. The social construction of teen pregnancy as a *growing* problem contributed to political support for policies such as family planning and has been critical in debates over the consequences of family planning policies, even when the statistics were questionable (Vinovskis 1988). Feminism also contributed to changing attitudes towards family planning policies, but the paradox for social scientists is that demographic trends did not affect perceptions of the levels of teen pregnancy. Academic researchers on teen pregnancy have recognized the incongruity that the definition of teen pregnancy as a social problem coincided with a decrease in birth rates (e.g., Furstenberg 1991). Still, gross numbers (for example, total births to teen mothers) created the popular perception of a crisis. Statistics help define perceptions of social realities and possibilities. Starr (1987: 54) has noted,

> An average is not just a number; it often becomes a standard. . . . Many regularly reported social and economic indicators have instantly recognizable normative content. The numbers do not provide strictly factual information. Since the frameworks of normative judgment are so widely shared, the numbers are tantamount to a verdict.

The existence and frequent public reporting of teen pregnancy statistics by themselves created public debate that led to policies attempting to limit teen pregnancies. Much other public reporting of statistics likewise shapes public debate: Newspapers and broadcast news regularly report unemployment and inflation figures, crime rates, and school test scores.

The distinction between practice and political legacies of statistics is useful in explaining

why accountability practices are so popular and what the potential consequences of the most commonly-discussed accountability systems might be in the long term for school politics. Most critics of high-stakes standardized testing point to the practice legacy, the way that high-stakes testing may narrow the focus of teaching and provide perverse incentives within schools and school systems. However, the political legacy is as important as, and in some important ways dovetails with, the practice legacy. High-stakes testing narrows how we judge schools as institutions and whose school success is important. Moreover, opponents of high-stakes testing rarely consider the political legacy of proposed alternatives. The most prominent alternative vision of accountability revolves around the outdated model of ascendant professionalism. A consideration of accountability's political legacy would require different alternatives to high-stakes testing, ones that would cultivate deliberate political connections between schools and communities.

Table of Contents

- The Importance of Political Legacies
- The Popularity of School Accountability
- Unexamined Assumptions of Accountability
- The Political Costs of Accountability
- The Political Weaknesses of Professionalism
- The Ground We Stand on
- Where To Go
- References (uses a second browser window)

The Importance of Political Legacies

I choose the term *political legacy* for statistics because statistical systems constitute a special example of how public policy creates long-term consequences for public debate. Those who study government from a variety of disciplines recognize that public policies set in motion political dynamics that shape the contours (and sometimes define the limits) of accepted political debate. Two parts of the original Social Security Act of 1935, pension insurance and Aid to Dependent Children (the federal program most call welfare), demonstrate the way that policies can define the political landscape. The pension insurance part of Social Security is a universal program; anyone who pays into Social Security as a wage-earner (as well as a beneficiary defined by law) is eligible for payments when older. The universality of the Social Security pension has made its basic features unassailable politically. By contrast, federal welfare was a means-tested program. Only poor people (and not all poor people) were ever eligible for federally-supported welfare programs. Unlike Social Security pension insurance, welfare was politically vulnerable because of its means testing. Since most people would like to live long, they think of Social Security as an important safety net. But most people do not want to be poor and, as critically, may not think they ever will be poor enough to be on welfare. The universality of Social Security has protected it politically. Thus, when President Ronald Reagan suggested changing the pension program in the early 1980s, politicians rallied to support the system. However, without universality, federal welfare had a much less powerful base of support, and the Republican Congress and President Bill Clinton ended the federal welfare guarantee in 1996. The original outlines of the two programs shaped future debate over them (Skocpol 1991).

The different histories of school desegregation in the South and elsewhere since 1954 are also results of a political legacy. The fundamental paradox of desegregation is that the South (including border states) had the most integrated schools in the country by the late 1980s (Orfield 1993). Southern schools have been more integrated because of two policies vigorously pursued by white, racist politicians and officials before 1954: state laws mandating segregation and policies of school and government consolidation. Because state law and intentional acts by school officials were an obvious cause of school segregation,

federal courts after 1954 had clear and convincing evidence of unconstitutional segregation in Southern systems and were willing to order far-reaching remedies in the late 1960s and early 1970s. In addition, Southern school systems are usually much larger than systems in many other states because of consistent success in consolidating school systems this century. For example, Mecklenburg County, North Carolina, has one school administration, so the suburbs of Charlotte are in the same school system as the city. In contrast, the suburbs of Boston are in school systems separate from the central city. Desegregation advocates in the South had two advantages stemming from consolidation. First, courts were more willing to order metropolitan desegregation plans in the South, after the *Milliken v. Bradley* (1974) decision required that judges find specific evidence of discriminatory intent to remedy metropolitan segregation in fragmented urban areas. Second, large systems made white flight more difficult. Because the South had both a history of state-directed discrimination and also large school systems, desegregation efforts in the region in the late 1960s and early 1970s were more vigorous and far-reaching than in the rest of the U.S (Douglas 1995; Orfield, Eaton, and the Harvard Project on School Desegregation 1996). The political legacy of statutory segregation and school consolidation made extensive desegregation more feasible in the South.

These stories, of government pension and welfare programs in one case and desegregation in the other, demonstrate the relationship between the structure of public policy and later political decision-making. To be sure, that influence is not one-way. A government is not an empty vessel easily manipulated by electoral and other political forces. Instead, government agencies have their own interests, and officials often act in their organizational interests (Balogh 1991b; Galambos 1970). Schools, like other public bodies, have their own professional and organization dynamics that mediate, rather than automatically reflect, outside influences. Thus, when we speak of a political legacy of school policies (including statistical systems), that legacy is part of a larger negotiation over the role of public schools. Two facets of that constant bargaining are particularly relevant to understanding the current school accountability regime: the limits of educators' professional authority and the local nature of schooling. First, as explained in the next paragraph, school administrators have tried to claim both bureaucratic autonomy and public acknowledgement of expertise involved in running schools. They have been far more successful in the former task than in the latter. In addition, schooling is a local, public service. Local political control of schools, and the close watch that one can theoretically keep over such institutions, may be one reason why school administrators garnered autonomy earlier in this century. One can thus view statistical accountability systems as one way to resolve the dilemma between granting autonomy and authority to educators and keeping them under some political control.

The political legacy of statistical accountability systems is important because support for publicly controlled schools is fragile. School administrators deliberately built a set of bureaucratic institutions in the early twentieth century to buffer themselves politically, in part by claiming the need for autonomy to exercise professional judgment and wield their expertise (Tyack 1974; Tyack and Hansot 1982). That autonomy, and the justification for publicly controlled schooling, has been on the wane since mid-century for several reasons. First, the civil rights movement targeted schools as one public institution that was treating poor and minority children unequally. The attack on school inequalities undermined support both from those who thought that inequality is morally wrong *and also* from those who had relied on state and local control of education to preserve bastions of private privilege (Kozol 1991). Second, the credibility of public institutions as a whole has deteriorated. In part, the Vietnam War and Watergate created a credibility gap between what public leaders said and what most citizens saw happening (Schell 1975); in addition, the internal politics of public agencies have damaged their ability to wield professional consensus as a political force (Balogh 1991a). Third, schools have been the target for half a century of accusations of ineffectiveness and soft standards. All of these events undermined the legitimacy of school administrators as autonomous professionals and public schools as worthy of financial

and political support (Tyack and Hansot 1982). Privatization, through charter schools or vouchers, represents one potential result of declining support for school systems as publicly financed and controlled organizations. The political legacy of current educational reforms, including growing development of statistical accountability systems, will define in some measure the future debates about schooling.

Return to Table of Contents

The Popularity of School Accountability

The public judging of schools by test scores is relatively new in the United States. School statistics have existed since the late 19th century, and claims to objective measurement of student achievement from the turn of the 20th, but achievement scores have typically been only for internal consumption *within* school bureaucracies until recently. In the wave of school criticism after World War II, ideological debates over progressive education and the needs of the Cold War were the explicit points of conflict; statistical evaluations were invisible in the 1940s and 1950s debates over schooling (Ravitch 1983: 71-80, 228-32; Spring 1989: 10-33). The public debate over Scholastic Aptitude Test (SAT) score trends did not exist until the mid-1970s, even though the decline in mean scores began in the early 1960s. The *New York Times* , for example, did not start reporting SAT scores annually until 1976 (Maeroff 1976). No network news broadcasts between 1968 (when the Vanderbilt Television News Archive began recording and indexing network news) and 1974 reported test scores as the substance of the story; the first networks to do so after 1967 were ABC and CBS on October 28, 1975. (Note 3. Uses second browser window.) The popular reporting of periodic student data, therefore, is of relatively recent vintage. One may consider statistics as one of many types of evidence and reasoning in public debate, such as the following list (meant to be an illustrative rather than a comprehensive typology):

Ideology
: Debates can focus on the purposes of schools and the perspectives offered in the curriculum or in teaching techniques. The attack on what progressivism had become by the 1940s is an example of ideological debate, as was the attack on outcome-based education in the early 1990s in Pennsylvania and elsewhere.

Representative Story
: Debates can center on real or apocryphal stories about education that represent the issue at hand. Anecdotes about high school graduates who cannot read (and the argued need for higher graduation standards) are an example of argumentation from representative story.

Statistics
: Debates over the quality of education in the 1980s, following the *Nation at Risk* report (National Commission on Excellence in Education 1983), are an example of discussion focused on statistics.

Direct Observation
: Debates can also focus on what individuals have seen, first-hand, in schools. I do not know of any national debate relying on directly observed evidence.

The self-evident explanation of the last statement suggests, in part, that we focus on statistics because having a "national discussion" based on personal, direct observation of schools is a contradiction in terms: we cannot each observe the nation's schools, and our judgment of "the nation's schools" will depend on second- or third-hand information. Still, most discussion of schools, and even school statistics, is local. Only thirteen network news broadcasts in the twenty-year period 1968-1987 reported statistical test score trends. (Note 3. Uses second brower window.) Most reporting on education, and most of what individuals

hear and read from popular media sources, is still in local news broadcasts and local newspapers. Why, then, have local educational debates generally assumed the importance of statistics, something that makes more sense for a national debate?

The common use of statistical mechanisms to gauge school effectiveness, including the power of standardized test scores, owes its existence to the tension between the development of a national debate over education in the twentieth century and the continuation of local decision-making. The result is a set of themes which dominates discussion in cities and states across the country and that borrow much of their character and assumptions from the national debate. In many cities and towns, for example, newspapers and local news broadcasts describe similar issues such as discipline problems and whether high school graduates are ready for the workplace. Several changes in schooling since the early 19th century have encouraged a national debate. First, educational reformers have typically borrowed from each other's ideas, spreading them from region to region. Second, professional educators and muckraking journalists in the late 19th and early 20th century explicitly campaigned in nationally-distributed journals against school corruption and the decrepit conditions in urban schools, on the one hand, and for professional autonomy on the other. Their campaign nationalized the Progressive Era education debate. Third, administrative progressives (as David Tyack has termed them) were successful in creating standard institutional routines in the first half of the 20th century, so that many school experiences adults remember now *are* much more similar across the country than adult memories of childhood were 150 years ago. We thus have a common set of experiences nationally, making the terms of debate familiar. Finally, the nationalization of politics more generally after World War II encouraged the debate over Cold War schooling described earlier. The civil rights movement and desegregation consolidated that national framework for discussion.

Still, the national educational discussion is a layer on top of and filtering down through older, local politics of schooling. Localism has remained a powerful force. It has controlled the politics of local *and federal* educational programs. For example, Southern members of Congress were critical in supporting federal vocational education programs early in the century because the federal government allowed Southern states to distribute funds disproportionately to white vocational programs and create different curriculum programs by race. The result was that vocational education programs served to reinforce the Southern caste structure (Werum 1997). Traditional federal deference to state action also modified and limited Title VI of the Civil Rights Act of 1964, whose implementation still helped force school desegregation in the South (Orfield 1969). Opposition to federal intrusion has limited national action to the present, including President Clinton's desire for tests created and organized by the federal government. Politicians are willing for schools to buy textbooks from national publishers, accepting a tacit national curriculum (Miller 1997). Federal *government* decision-making, however, threatens more than local control of curriculum; it threatens local political networks and ways of doing business. Local political control of school policies and funding thus vie with the national debate. The result is frequently a set of variations on common practices, resulting in the illusion of local control in many school matters. Standardized testing and accountability systems are one example of that limited variation. States are free to choose commercial tests, develop their own, or not to engage in high-stakes testing at all. Today, however, most local school systems or states test children in the spring using multiple-choice tests with scores that schools can compare (using the publisher's data) against a norming population of children in the same grade. In the past dec de, many states and local districts have added real consequences for the tests, including publicly releasing score data. The result is a patchwork of high-stakes testing that covers most of the nation. Despite theoretical local choice about standardized testing, one way of publicly judging schools has become dominant.

The emergence of contemporary school "accountability" dependent on test score results combined an existing set of practices (standardized testing) with the judgment of local schools within a national framework. Within a decade, public judgment of schools by test

statistics became common, after the College Board publicized the decline in mean SAT scores, states began instituting minimum competency tests, and the National Commission on Excellence in Education published *A Nation at Risk* in 1983. Two historical perspectives underline the importance of understanding the political implications of school accountability systems.

- Accountability has turned the use of educational statistics upside-down. Statistics bolstered the claims of administrators to expertise early in this century, but politicians and popular news media now use statistics to judge school systems. This reversal shows the weakness of local school administrators in claiming professional authority. Autonomy within bureaucratic organization, not public respect of their expertise, is the primary power of school officials.

- The popularity of published test scores obscures alternative ways of judging schools. In less than twenty-five years, statistical accountability has become so ubiquitous that it appears inevitable. The change has been, in retrospect, both breathtaking and alarming in its speed. Political debate over the meaning of statistics has largely eclipsed other ways of describing what happens in classrooms.

The dominance of educational test scores today hides the fact that we did not have to use statistics as the dominant way of describing schools and their problems, and that in the past we have used many other means. Even when we evaluate local schools using nation-wide questions, we can use many sources of information. Assuming we must use primarily statistics is dangerous. We must remember that the evaluation of schools by test score statistics is one *among many* possible ways of seeing education through both national and local perspectives. Whether we made that choice consciously or wisely is a different question.

Return to Table of Contents

Unexamined Assumptions of Accountability

One consequence of public policy is the definition of legitimate debate and, by extension, what is not part of mainstream public discussion. Often, the assumed axioms underlying policies silence other relevant concerns (Fine 1991: 32-34). Despite more than twenty years of debate about the statistical performance of students in the U.S. and the proper direction for school reform, remarkably few voices in public have questioned the primary assumptions behind the move towards accountability. This silencing shows what we are avoiding when we speak glibly of a political consensus around school accountability. While we are agreeing to high-stakes testing, what uncomfortable issues are we not discussing? The broad political legacy of statistical accountability systems is the narrowing of legitimate topics for public debate. We do not often discuss the purpose of accountability or who will be making the key decisions to keep schools accountable.

Accountability for what purposes?

The dominant discussion of accountability leaves vague the goal of accountability mechanisms. The improvement of schools is an insufficient goal because accountability is fundamentally a political and not a technical process. Accountability has multiple meanings, in both a general sense and also the current sense in education of statistical judgment (Darling-Hammond and Ascher 1991). The apparent consensus for "accountability" hides the differences (and the conflicts) among the following meanings of statistical systems.

Judging public schools as institutions . One may use test score statistics to judge schools as a set of institutions. This sense of accountability (judging the worth of schools in

general by test scores) is one of the most widely used tools in school politics. The annual release of average SAT scores in the late 1970s prepared the ground politically for the claim of declining school effectiveness made by the National Commission on Educational Excellence (1983). One political legacy of judging public schooling by test scores is the assumption that schooling is a monolithic entity that fails or succeeds as a single body. What this myth of a monolithic system hides is wide variations in schooling, especially between poor and wealthy schools (Kozol 1991). Another political legacy is that, after intense media focus on statistics that suggest poor schooling, citizens may face difficulty reconciling popular conceptions of failing schools with information gathered in other ways. Polls consistently show that parents' perceptions of their local schools are more positive than their perceptions of schooling nationwide (e.g., Rose, Gallup, and Elam 1997). In addition, private interests may subvert policies based on the gross judgment of schools. For example, some wealthy parents in one Michigan district deliberately pulled their children out of high-stakes standardized testing when they perceived that it might hurt their children (Johnston 1997). They may well have been willing to have high-stakes testing for "other people's children" (to borrow from Lisa Delpit's 1995 book title) but not theirs. This consequence is the educational equivalent of urban development NIMBY (Not in My Back Yard) syndrome.

Judging teachers and other educators . One may also justify accountability as a way to raise (or clarify) expectations and goals for teachers and administrators. An explicit part of accountability systems in the last few years has been the evaluation of teachers, principals, and other administrators. For example, the Tennessee Value-Added Assessment System, passed in 1992, originally mandated statistical measures of student gain as part of personnel evaluation (Educational Improvement Act of 1992). An earlier variant of judging teachers, schools, and school systems by comparative statistics was the U.S. Department of Education's "Wall Chart" instituted by Terrence Bell as an attempt to spur reform (Ginsburg, Noell, and Plisko 1988). This use of accountability, focusing on teachers and administrators, is the one most criticized as encouraging teaching to the test and "gaming" test results (Cannell 1989; Glass 1990; Madaus 1988, 1991; McGill-Franzen and Allington 1993; Merrow 1997; Shepard 1991; Smith 1991; Smith and Rottenberg 1991). The political legacy, however, may be even more harmful: By setting up a system based on the distrust of teachers, we make alternative ways of judging teachers and schools more difficult (Fisher 1996; Sizer 1992: 188-89).

Judging students . In many states and school systems, standardized tests have high stakes not only for educators but also for individuals students, as scores can be among the criteria for entrance to academic programs, grade promotion, or other real rewards and punishments in schooling. The use of tests to sort students U.S. began with monitorial schools in the early nineteenth century and admissions tests to early public high schools (Kaestle 1973; Labaree 1988; Reese 1995). More recently, the use of so-called minimum competency tests emerged in the late 1970s as a response to allegedly lowered standards of public schools (Bracey 1995b). The rationale of using tests to make students accountable is that, having test scores as a clear goal, students and schools would meet the expectations (Ravitch 1995). One potential legacy of such high stakes, however, is the rhetorical scapegoating of students. Calhoun (1973: 70-72) describes one purpose of testing in schools as displacing blame for ineffective teaching onto students. If a student fails a test, one may reason, the failure is the student's intelligence and lack of diligence. That consequence is already evident in many states with high-stakes testing. In Tennessee, for example, the teachers union pressed to exempt scores of students with disabilities from teacher value-added statistics ("Sanders model to measure 'value added'" 1991). One might presume that children with disabilities are those on whom we should *most* focus attention in evaluating teaching effectiveness. Yet teachers asked for the exclusion of scores because, the union argued, including such scores would be unfair to teachers. The displacement of blame for failed schooling onto students is a legacy of testing that existed well before high-stakes standardized testing, but accountability systems may exacerbate such tendencies (e.g.,

McGill-Franzen and Allington 1993; McGrew, Vanderwood, Thurlow, and Ysseldyke 1995; National Center on Educational Outcomes 1994).

Judging public policy . One might use standardized test scores (like other information) to evaluate public policies. The National Assessment of Educational Progress (NAEP) tests, begun in 1969, is theoretically a means for using non-high-stakes testing to evaluate public school policy with objective data. NAEP data is at the heart of some recent debate about school and student performance (see Berliner and Biddle 1995, 1996; Stedman 1996a, 1996b). However, demands to use the NAEP to judge educators and students in high-stakes systems is threatening to compromise NAEP's use as a lower stakes way to gather information about student performance (Jones 1996; Koretz 1992a). One problem is the technical and fiscal demands of high-stakes versus low-stakes systems. In addition, however, is the ideological debate about the use of information. Can one maintain a low-stakes statistical system in the face of political pressures for high-stakes accountability?

Building organizations . In a broad sense, standardized testing supports the determination or control of curriculum content at the state and national levels. Some such as Ravitch (1995) explicitly advocate curriculum content standards and see teaching to the test as valid with appropriate testing and content. One consequence of statistical accountability, however, is the creation of new public and private organizations producing educational statistics. Publicly, states now have accountability or evaluation offices whose job is to provide the technical expertise in analyzing test data, and the federal government has the National Center for Educational Statistics, which contracts out NAEP as well as compiling and disseminating a wide variety of educational statistics. Private organizations supported by testing are the companies that write and sell tests or contract with agencies for the creation of specific tests. With each public release of test score statistics, popular news sources, politicians, administrators, and the public rely more on relatively anonymous technocrats to explain what is happening in schools. Other new professions this century, such as nuclear science, have also staked their claim to expertise on political factors (Balogh 1991a). The fact that this reliance on statisticians stems from political pressure for school reform usually escapes notice.

Marketing . Schools occasionally use student statistics as part of public marketing strategies, either to attract students who have choices (as in selective colleges) or to bolster public support. One of the largest metropolitan school systems in the country recently produced a pamphlet boldly titled, "Our Students' Test Scores Reflect Academic Achievement" (Hillsborough County Public Schools 1997). While one paragraph cautions that test scores are not the sole basis for evaluating students or schools, the rest of the pamphlet trumpets above-average achievement. Public relations was a strong motivation behind what Cannell (1989) called the "Lake Wobegon" effect of claiming high test scores in public reporting through the use of outdated norms. The use of accountability data for marketing is an open secret among administrators. As Dennie Wolf said in the John Merrow documentary *Testing . . . Testing . . . Testing* (1997), "Districts sell real estate based on test scores." With the decline of administrative authority described elsewhere in this article, superintendents have considerable interest in boasting about their systems using any tools at their command.

These varied purposes of accountability are not necessarily congruent. The use of test scores to bash public schools is not compatible with a nuanced debate over public policy, and students and teachers may have conflicts of interest when tests have high stakes for both. In addition to inconsistent purposes, the aims of accountability do not easily include other issues relevant to education: equity, the direction of curriculum, or the purposes of education more broadly in a changing world (Darling-Hammond 1992). One dominant assumption of accountability systems is that the goals of education are agreed upon and we need only establish a system to measure whether schools and students meet those goals. The creation of statistical accountability systems may freeze the assumption of a single purpose

of statistical accountability into a framework for the politically accepted discussion in education for years hence.

Who keeps schools accountable?

A second unexamined assumption is that central bureaucracies and popular news media are the logical, natural places for holding schools accountable for performance. In most school testing regimes, central offices (at the state or local level) are responsible for the general logistics of testing and compiling results. Results at some level are then available to administrators, public boards of education, and media organizations. In many states and regions, newspapers publish test score statistics, often ranking schools or systems based on the scores. But who is *not* among the direct targets of test score dissemination is as important as who *is* .

Judges and advocates monitoring school system compliance in discrimination cases . Judges and advocates overseeing compliance with nondiscrimination orders (such as desegregation) generally are not intended users of "accountability" information. Despite promises by school systems to pay closer attention to achievement in desegregation cases, local systems have a very spotty record in demonstrating success after the end of desegregation orders. Orfield, Eaton, and the Harvard Project on School Desegregation (1996) has compiled evidence that, in several of the major cases this past decade, school districts released from desegregation monitoring by the courts not only experienced resegregation but growing achievement gaps between white and minority students. The new accountability system does not appear geared to keep systems accountable in this respect. Many advocates appointed to monitoring and advisory commissions have reported to Orfield and his associates that local systems have either denied information (such as disaggregated test scores) outright or made the gathering of data extremely difficult. In addition, the Supreme Court decision in *Missouri v. Jenkins* (1995) declared that district court judges should consider test scores as marginally important (at most) as a measure of compliance with racial equity requirements. The only major case where a court has continued to monitor standardized test scores as part of a major equity lawsuit has been in New Jersey, where the state's supreme court continues to criticize inequalities between the education offered children in the wealthiest and poorest systems of the state (*Abbott v. Burke* 1997). In the past five years, the court has broadened its focus from just monetary support of schools to include measurable outcomes. The New Jersey Supreme Court has been a lonely exception to the general rule, especially in the federal judiciary: Accountability does not appear to require even reasonably equitable outcomes.

Parents and the general public . Parents receive test scores of their children, but rarely do they or the general public have direct access to test score results *or their limitations* . Popular news sources (television, radio, and newspapers) mediate the transmission of information, often deleting information critical to understanding the limits of such data or transforming the statistics in ways either incomprehensible to readers or to create invalid statistical comparisons. The reporting of high-stakes test data by Nashville metropolitan newspapers form a case in point. Beginning in 1993, the state of Tennessee reported test results of schools and districts using a complex statistical system called the Tennessee Value Added Assessment System. The state's newspapers have quickly rushed to print school-by-school scores including rankings, even where schools many rankings apart had negligible differences in scores (in other words, when the rankings were unjustified by the statistics). For example, in 1996 the Nashville *Tennessean* transformed the value-added scores into percentile ranking, even though the technical documentation for value-added scores would not support such an interpretation (Bock and Wolfe 1996: Chaps. 5-6; Klausnitzer 1996; Tennessee Department of Education 1996). Why did the *Tennessean* transform value-added scores that were the result of a prior statistical manipulation, and why did the paper then rank schools? One reporter explained:

> We chose to report in percentile ranks because it helps people see how their school stacks up against the rest of the state, and because this information is not available anywhere else. It was calculated by *The Tennessean* ... [because] we wanted to offer something unique. We also wanted to answer our readers' number one question about the test scores: How does my child's school compare to the other schools? (Lisa Green, e-mail to author, December 5, 1996)

In addition, the newspaper reported percentile rankings by tenths (for example, 50.1 instead of 50th percentile). The same reporter acknowledged that the newspaper staff did not consciously justify that apparent precision:

> There's really no need to report these numbers down to the tenth of a percentile. However, the programming for the site was written last year ... so the computer automatically included the decimal place, and we didn't think it was necessary to take it off. (Lisa Green, e-mail to author, December 5, 1996)

In this case, a metropolitan newspaper's desire to have "something unique" conflicted with its readership's interest in having clearly understandable information to interpret independently, or even information with a justifiable level of detail. Even if one assumes that the value-added scores are comprehensible, transforming those into percentile rankings was neither valid nor necessary for rankings (itself a method of reporting scores which the state's external evaluators recommended against). In no case did the newspaper note what the evaluators clearly stated: that school scores were unstable and could not be relied on for clear distinctions in performance (Bock and Wolfe 1996: Chap. 5-6). The dissemination of information through two intermediaries (the state government and news sources) in essence created one dominant way to analyze scores in the metropolitan Nashville area: how did schools "stack up" in competition with each other? The false precision in percentile rankings suggested that readers could rely on the numbers as rigorous, objective facts. The accuracy of newspaper reporting is also questionable; the *Tennessean* had to reprint its comparative tables in 1994 because of acknowledged gross errors in reporting ("How Midstate Schools Stack Up" 1994a, 1994b). While comparisons among schools may be appropriate in some ways, the presentation of school scores suggested a certainty which was incompatible either with the statistical calculations or the mediation of state agencies and newspapers in transmitting test scores.

Moreover, the dissemination and discussion of today's school accountability systems strip parents and the general public of control and ownership of information. In the case of Nashville, a reporter reduced parental evaluation of schools to examining rankings in a table, akin to sports league rankings (see Wilson 1996). One might contrast the typical method of disseminating accountability statistics with two alternative local methods of accountability: the "visiting committee" of town elders in the eighteenth and early nineteenth-century district schools, on the one hand, and the calculation of dropout statistics by a Hispanic activist organization in Chicago in the 1980s, on the other. In many district schools, a small committee of citizens held the power of hiring and firing over schoolteachers and could visit the school at any time (e.g., Cohen 1973: 407). Accountability in district schools was a rough-and-tumble affair, often unfair to teachers, but local citizens could form judgments in a simple way: watching classrooms. Independent gathering of data today is also possible. In the 1980s, Aspira, Inc., a Hispanic activist organization, suspected that official dropout statistics from the Chicago public schools were inaccurate or fraudulent and conducted its own research. Activists then used the independent statistics to help prod Chicago towards urban school reform (Hess 1991: 7-21; Kyle and Kantowicz 1991). In both cases, individuals at the local level produced and acted on their own judgments of schools. Reliance on centrally-calculated statistics in accountability systems often overrides local, independent judgment of schools.

The fundamental issue of control is directly connected to the purposes of accountability: Individuals in different roles would ask different questions of accountability mechanisms. Politicians might ask whether schools "measure up" to some standard (such as a national norm). Business leaders might ask about workplace-related skills and behavior. College faculty would want students to have some intellectual foundation. Parents might ask whether their children are getting enough individual attention. Who should be asking the hard questions about schools? The history of the Common Core of Data (a set of education data collected by the federal government since the early 1970s) illustrates the difficulties of creating an explicit consensus. Because of pressures within government, doubts about its utility and cost, and disagreements about what it should measure, the Common Core of Data for many years gathered relatively innocuous information in a history Janet Weiss and Judith Gruber (1987) described as "managed irrelevance." Of all the information used by the National Commission on Excellence in Education (1983) to lambaste the condition of schools, none came from the official federal education database (Weiss and Gruber 1987: 370). What we face is not an explicit consensus but a hidden one, never debated clearly, founded on the spread of standardized test scores. Statistical accountability systems suggest an objectivity and universality of coverage which is impossible. As Sizer (1995: 34) noted with regard to the debate about educational standards, "The word *system* has come up again; . . . Essentially, it implies a technocratic approach." We should not evade the political question of the purposes of schools through the production of statistics. The current penchant for statistical accountability systems diverts resources to a mechanism that hinders discussing the nuts and bolts of schooling. We hide behind the apparently objective notion of an accountability system.

Return to Table of Contents

The Political Costs of Accountability

The political legacy of statistical accountability systems is complex because of the different possible aims of (and justifications for) accountability and also because statistical systems will vary among different states and districts. Nonetheless, one can identify several broad patterns which stem at least in part from the proliferation of statistical accountability systems. Two legacies have seriously damaged our collective ability to have reasoned, broad discussion about the aims of schooling and reasonable public policy. Statistical judgment of school has narrowed the basis on which we judge schools and has also encouraged impatience with school reform.

Narrowed Judgment of Schools

Technocratic models of school reform threaten to turn accountability into a narrow, mechanistic discussion based on numbers far removed from the gritty reality of classrooms. Over the past twenty years, the dominant method of discussing the worth of schools in general has been the public reporting of aggregate standardized test score results. Popular news sources typically distort and oversimplify such findings (Berliner and Biddle 1995; Darling-Hammond 1992; Koretz 1992b; Koretz and Diebert 1993; Shepard 1991). The recent public debate over schools is not rich, reliant on multiple sources, or nuanced. Nor is the reliance on statistics inevitable in national discourse, despite recent history. Prior waves of reform, such as concerns about math and science education in the 1940s and 1950s (whether one agrees with their goals or not) did not need test score data as motivation or evidence (Ravitch 1983).

Test-score data and its use have pushed other issues to the margins. The aftermath of the 1983 report *A Nation at Risk* eclipsed two major policy initiatives of the first Reagan administration. The early 1980s saw dramatic cutbacks in the support of the federal

government for state and local public schools. At the same time, social conservatives both in and out of the Reagan White House were arguing for the creation of vouchers to support parents sending their children to private schools. Neither of these issues, however, were part of the central discussion of education policy after the release of *A Nation at Risk* . The dominant discussion in popular news media revolved instead around declining test scores, the presumed responsibility of schools for national economic decline, and how to tighten academic standards (Berliner and Biddle 1995; Bracey 1995b). Few mentioned changes in the federal budget or privatization proposals, even though one was a concrete policy of the Reagan administration and the other was a radical proposal for changing the governance of schools. Ironically, the dominant discussion suppressed issues which concerned both liberals (upset at budget priorities) and social conservatives (wanting vouchers).

More recently, New Jersey Governor Christine Todd Whitman tried to argue that a standards-based accountability system alone could improve the state's schools. Her department of education responded to the state Supreme Court's call for equity with state-level achievement standards but no added resources, despite the state's history of vividly unequal funding among school systems. The argument by the executive branch was that standards, by themselves and despite existing funding inequities, would create school improvement. The assumption by Whitman is that test-based school accountability, as a technocratic mechanism with threatened sanctions, is sufficient to change schools, even schools with the worst records. The state court agreed with the governor in that New Jersey could have state-level standards but disagreed with the argument that funding was irrelevant. It then ordered the state to improve its funding of poor schools (once again) (*Abbott v. Burke* 1997). New Jersey is fortunate in having one branch of government able and willing to articulate a complex view of what school reform requires. In general, however, extending public discussion of schools beyond test-score statistics is difficult.

Impatience with Reform

On a political level, impatience with reform and the cyclical reporting of statistics encourages the dominant myth of contemporary educational politics, that schools continue to decline in quality. (Note 4. Uses second browser window.) That myth encourages a cynicism towards reform strategies. We should not be surprised that we have witnessed several "waves" of reforms since the regular publishing of SAT scores began in the 1970s. The mundane details of statistical accountability systems encourages fads. Without a concrete sense of what children and teachers should be or are doing, the public compares statistics against a set of arbitrary benchmarks.

On a practical level, statistical accountability produces both undue impatience with reform and laxity towards incompetence. The yearly reporting of test scores creates an artificial schedule for judging schools: Do they improve by the next set of annual tests? The periodic nature of reporting school statistics drives the disposal of reform writ large, because policy changes cannot change classroom practices on a deep and fundamental level or become institutionalized in a short time (Lipsky 1980; Tyack and Cuban 1995). Yet, paradoxically, the annual time-frame of standardized testing gives too *much* time for weak teachers to flounder without guidance or correction. Pinning personnel practices to annual testing may undermine the obligation of fellow teachers and administrators to keep a close eye on teachers without the necessary classroom skills. Principals may feel inclined to give poor teachers until the following cycle of annual tests to improve. For children, however, a year of being with an incompetent teacher can be extremely destructive. The problem is in part one of inappropriate time scales. Annual tests are too infrequent for appropriate guidance of instruction or evaluation of teaching, while they are too frequent to measure broader changes in schools.

In addition, standardized test accountability discourages the evaluation of what happens in the classroom. As long as a school or teacher has adequate test scores, what happens in the classroom is irrelevant. Similarly, poor test scores indicate needed change, no matter what happens in the classroom. The philosophy behind such practice-blind evaluation is putatively to give teachers autonomy. As the designer of one state's accountability system explained, accountability statistics allow teachers to make their own choices (Sanders and Horn 1994). Ultimately, however, this diminution of practice undermines teacher and school power, for several reasons. First, teachers do not usually have time to review and evaluate on their own a wide array of alternative teaching methods; they need support in selecting, adapting, and implementing different methods and curricula. Second, parents and other citizens *do* care about what happens in classrooms. Schools trying dramatic departures from normal practices face (sometimes very reasonable) criticism from parents even when the intent is to respond to the accountability system. Separating accountability from the sense of what a "real" school is (Tyack and Cuban 1995) is deceptive in the long run. It gives schools the following message: "Make your choices because we only care about test statistics. But we won't give you enough support to follow up on your choices, and in the end we will condemn your choices if they violate our ideas of what schools should be." One consequence of statistics-driven impatience is increased cynicism among teachers and administrators and their uncertainty about what the public really wants. Discussions isolated from what happens in schools may be politically alluring and attractive to popular news sources, but test scores drive a wedge between schools and the students and public they serve.

Parallels between Practice and Political Legacies

The political legacies of high-stakes statistical accountability systems parallel the practice legacies in two respects. First, narrowed political judgment of schools is the macropolitical equivalent of teaching to the test, a narrowing of the curriculum. Researchers have documented the tendency for teachers to narrow their focus to content and styles which they perceive will result in high test scores (Madaus 1988, 1991; Smith 1991; Smith and Rottenberg 1991; Shepard 1991). Relatively few teachers, faced with the onslaught of standardized testing, are willing to innovate. Meier (1997: 9) writes,

> The danger here is that we will cramp the needed innovations [in teaching] with over-ambitious accountability demands. Practical realism must prevail. Changes in the daily conduct of schooling . . . are hard, slow, and above all immensely time-consuming; they require qualities of trust and patience that we are not accustomed to.

High-stakes accountability is not a system that demonstrates trust in teacher's capacities. By signaling massive distrust, high-stakes testing instead provides low expectations for teachers (Sizer 1992: 110-13). Imagine the result of a thought experiment: the plight of John Dewey's University Lab School teachers under a high-stakes system. One might like to spend an extended time exploring history and science through the concrete example of textile manufacturing (Dewey 1899). In a modern accountability system, however, the state will test the children in March or April, with much of the test based on several dozen discrete skills. Whether the children can understand the role of textile mills in 19th century economic changes, or whether they can explain what principles allow a loom to work, is irrelevant to accountability systems based on standardized tests. Balancing such competing demands is extremely difficult. Teachers and schools who fight the pedagogical consequences of high-stakes testing are relatively unusual. Whether one agrees with the appropriateness of multidisciplinary teaching for some or all children, one cannot confuse the expectations of today's statistical accountability systems with expecting children to understand connections between what they see in life and academic disciplines. The latter is of a higher order of magnitude entirely. Relying on standardized tests and high-stakes production of test statistics is itself a dumbing-down of political debate and expectations for schools.

Similarly, impatience with reform and fad fetishes are the macropolitical equivalent of being impatient with children's progress. The aggregation of test score data often gives teachers and administrators incentives to exclude students whom they feel will harm test figures. Repeated reports of test scandals, the plea by teachers in Tennessee to exclude students with disabilities from their statistics, and variations in the proportion of students tested provide continuing evidence of the perverse incentives high-stakes testing provides (Glass 1990; Madaus 1988, 1991; McGill-Franzen and Allington 1993; McGrew et al. 1995; Smith 1991; Smith and Rottenberg 1991; Shepard 1991). These incentives perpetuate a dynamic of educational triage, wherein those who have the best chance to survive in life because of other circumstances also have the best opportunities to learn (Fuchs and Fuchs 1995; Sapon-Shavin 1993).

Return to Table of Contents

The Political Weaknesses of Professionalism

If accountability based on standardized tests encourages a narrow political discussion about education and impatience with schools, alternatives proposed by critics of standardized testing confront the same history that engendered statistical accountability. Dissenters from the accountability "consensus" exist, from longstanding standardized testing critics at FairTest (http://www.fairtest.org) to the Coalition for Essential Schools (http://www.ces.brown.edu) to Teachers College professor Linda Darling-Hammond and Arthur Wise, current president of the National Council for Accreditation of Teacher Education (NCATE). Each opposes the idea of motivating school reform by standardized testing. The proposed alternative methods of motivating better teaching include performance (sometimes called authentic) assessment of students, peer evaluation of teaching, and either creating a second tier of high-status teachers or restricting entry into a limited number of high-status positions within teaching. Advocacy of greater professional authority in education have generally focused on teacher education and preparation (e.g., Darling-Hammond, Wise, and Klein 1995; Holmes Group 1986; also see Labaree 1992), but includes accountability; for example, Wise has been concerned with the deskilling of teachers since *Legislated Learning* (1979). In general, the critics of standardized testing seek greater teacher autonomy and respect from the public, and in that way we might call professionalism the central value of the dissenters (e.g., Darling-Hammond 1988, Haefele 1992). Wise and Leibbrand (1993: 135) write that, "Hallmarks of a profession include mastery of a body of knowledge and skills that lay people do not possess, autonomy in practice, and autonomy in setting standards for the field." If teachers could successfully professionalize, Wise and others suggest, they would gain more respect from the public and earn the autonomy needed to improve schooling (e.g., Wise 1994). The logic of professionalism is very appealing with the explicit parallels to the professionalism of medicine (Starr 1982). It links mechanisms within schooling (who controls decision-making) to the public status of teachers and the politics of schools. Professionalism appears to be politically astute.

Professionalism, however, is not likely to be a successful gambit in schooling, for several reasons. Most importantly, professional ideology is politically unpalatable in the late twentieth century. Trying to use professionalism misunderstands the historical context for the ideology of expertise and its widespread (political) success a century ago. Professionalism in the form of high-status, science-based occupations like medicine and engineering was one response to the chaos of industrialization and changing class structure (Wiebe 1967). Its early proponents argued that the complexities of modern life required technical expertise to solve public policy and practical problems. However, professions include more than high-status jobs, with occupations as diverse as architecture and craft work like plumbing. A profession typically involves three dimensions: a claim to specialized

expertise, some informal or formal credentialing to control entry into the occupation, and autonomy on the job (Friedson 1984). Classroom teaching falls partway among all three dimensions. Classroom teaching does involve some skills that few could walk in off the street with, but the general public has far more knowledge of what happens in classrooms (and is more willing to make second judgments of teaching) than fields like surgery. Long-term teaching requires credentials, but many school systems hire uncredentialed personnel on an emergency basis. Finally, public schools operate as loosely coupled organizations (Weick 1976): Most teachers can shut their doors in the face of some supervisory directives, but material conditions (such as the textbooks available) circumscribe their autonomy on the job, and they face other demands they cannot ignore, such as the official curriculum and standardized tests. We should see the ideology of professionalism thus as attempting to emulate a relatively small slice of all occupations with professional traits rather than, as is typically assumed, making teaching a "real" profession. Teaching already is a real profession, though one with less claim to specialized expertise and less autonomy than advocates of teacher professionalism would want.

Professionalism theories today appeal to an outdated ideal of insularity and ascendant authority. The worst excesses of school bureaucracies today stem from *successful* professionalism, albeit not in the classroom. Superintendents at the turn of the century argued that schools needed to be away from political battles that would harm the integrity of school systems. Creating an autonomous professional unit (a central school office) would improve administrative efficiency and rid schools of corruption (Tyack 1974; Tyack and Hansot 1982). Their success accelerated the bureaucratization of urban school systems. Today, however, professionalism is no longer unquestioned. School administration has credentialism and relative autonomy on the job, but not as much claim to specialized expertise as sixty or seventy years ago. Not only are North Americans far more skeptical of professional authority than fifty years ago (as discussed earlier), but capital mobility is impinging on professional authority in a wide range of fields. The parallels made between teacher professionalism and medical professionalism is jarring. One cannot today call medicine an autonomous profession when doctors are complaining that clerical workers and financial officers in health maintenance organizations are limiting their clinical decision-making (Bodenheimer 1996).

In addition to ignoring the historical decline of professionalism, arguments for advancing teacher professionalism undermines democratic control of schools. As Strike (1990: 362) noted, "Professionalism is nondemocratic in that it appeals to political values other than those of popular sovereignty to legitimate its authority." Peer review of teaching (e.g., Haefele 1992) is a case in point. Civil rights activists may not want teachers to have virtually unlimited autonomy in the classroom. Bob Peterson (1997: 4) explained, "A potential problem with the strictly professional union approach [to accountability] . . . in many urban districts has distinct racial overtones. Is peer evaluation the exclusive province of teachers and administrators or should parents and community members play a role?" Especially as the teaching force's demographics diverges from those of students and parents (Justiz and Kameen 1988), relying on professional-only evaluation may insult parents of a school who expect a role in school governance. Having an expertise-based evaluation system conflicts with U.S. traditions of democratic control, upon which civil rights activists have based advocacy of school governance councils. Some critics of standardized testing, such as Wilson (1996), point to British school inspections as an alternative to statistical accountability. The heart of the British inspection system, however, was until recently a self-perpetuating corporate body selected by and from experienced teachers. One may (as Wilson did) use school inspection to point out the problems in high-stakes accountability. One may not, however, successfully import the insular assumptions of professionalism to late 20th United States public schooling.

Professionalism is the dominant alternative to standardized-test-based accountability. Other critics of standardized testing-based accountability may not be as explicit as Wise in

their advocacy of professionalism, and they may not agree with his proposals to limit entry into high-status positions in teaching. Still, they argue for more decision-making power in the classroom and school and see the bureaucratization and centralization of authority as one of the reasons why standardized testing is flawed. Thus, Kenneth Peterson (1995: 4) argues that one of the key principles in teacher evaluation should be to "place the teacher at the center of evaluation activity." In that respect, the professionalism label is a useful heuristic device for understanding opposition to standardized testing. Despite its intriguing hypothesis (that status and autonomy are the key to educational reform), professionalism is unlikely to supplant high-stakes accountability because it is politically untenable.

Moreover, professionalism addresses primarily concerns inside schools (autonomy of teachers). Publicly, professionalism only changes the superficial aspect of teacher status, not the public dissatisfaction and disconnection which schools face more broadly. Several historical changes have fragmented what is supposedly a common public commitment to education. The aging of the population since the height of the baby boom has shrunk the political power of parents. In addition, the civil rights movement and a political coalition of fundamentalist Protestant organizations have stripped school officials of any broad political consensus. Finally, the fragmentation of urban politics and suburban growth has encouraged continued racial and class segregation (albeit in new forms), making common interests in broad school policies difficult (Katznelson and Weir 1985). While I doubt professionalism's proponents would ever claim that it is a panacea, they have nonetheless pinned their hopes for dramatic school reform on a model that would not solve the major problems of school politics today.

Return to Table of Contents

The Ground We Stand on

Like the expansion of Israeli settlements in occupied territories, the continuing spread of standardized testing has created "facts on the ground" which have transformed both schools and the politics of education. To ignore the educational landscape around us, or to wish it would go away, is unproductive. Those who disagree with the assumptions of high-stakes, testing-based accountability must acknowledge that standardized testing is likely to become even more prominent in the short-term. This understanding should not prevent advocates from fighting the trend where possible. Local victories against high-stakes testing are important both to the children involved and also as a standing alternative to technocratic accountability. Nevertheless, we should see clearly what is and is not possible in the near-term future.

The Future Growth of Standardized Testing

Standardized testing connected with high-stakes accountability systems is likely to become more prominent in the next five years in the majority of states. The Education Commission of the States (1997) recently reported that almost half of all states have implemented or are planning public accountability systems using statistical measures. Some additional states may use the national tests advocated by President Clinton (if the tests exist). Some like Tennessee will design their own accountability mechanisms. Others like New Jersey will create a set of content stand..ds ..ith the promise of new tests and accountability tied to the content standards. The federal government and states will then spend millions of dollars developing tests, field-testing them, and supporting their use. In the meantime, popular news sources will continue to report annually the average SAT scores and tests currently used in local jurisdictions. Within five to ten years, some states will begin the mandated use of exams replacing or supplanting current off-the-shelf commercial tests.

Moreover, the political debate over tests is likely to center around the federal relationship

between Washington and the states or (with privatization) public oversight of private schooling. For the duration of President Clinton's term, the administration is likely to support national tests, and governors who dissent (like Virginia's outgoing Governor George Allen) will do so not because they disagree with high-stakes tests but because they wish states to design their own independent standards. If federal courts, using *Agostini v. Felton* (1997), allow tuition voucher programs to proceed, state legislatures may contemplate mandatory use of high-stakes testing for private schools accepting public funds. The debate would then shift to public control of private educational institutions. A vision of the future debate may be *Ohio Association of Independent Schools v. Goff* (1996), in which a federal appeals-court panel concluded that Ohio's requirement to test private school students was constitutional. Those who disagree with all high-stakes testing will be at the margins of debate in the near future, except where they make alliances with others (as in the Congressional fight over national tests).

Limits on High-Stakes Testing

High-stakes testing has some significant weaknesses, despite the near-term growth we can expect. Some of the same dynamics which have limited the accountability use of performance-based, open-form testing will also shape standardized testing. Simply put, developing tests is expensive. The Tennessee legislature recently delayed the implementation of new subject tests for high school students to use in the value-added statistical system because, according to the bill's sponsor, the state could not afford the $10 million development cost (Educational Improvement Act Amendments 1997; Finn 1997). In addition, political adversaries may well use the management and pedagogical problems of new testing and accountability systems as a pawn in broader partisan battles. California's recent educational history is a case in point. Questions about the utility and propriety of performance-based tests combined with the expense of development and testing to kill the California Learning Assessment System. The governor, state superintendent, and legislature at the time were at odds over the purpose of the system, and that political conflict fed a controversy started by conservative critics over the ideological content of the tests, dooming the largest experiment in performance-based accountability to date (Kirst 1996; McDonnell 1997). Observers of merit pay have noted that political dynamics involving fairness and incentives to cheat typically kill merit pay systems (e.g., Glass 1990). The same may happen to the next generation of high-stakes accountability.

Contraction of the Meaning of "Public"

Despite the weaknesses of high-stakes testing, the short-term consequence of more standardized testing may be intensified criticism of public schooling and cynicism about the purposes of public educational systems. Schools need to be "public" in the sense of public involvement and political commitment (Fine 1991: Chap. 9; Katz 1992). However, the ranking of schools and teachers is inherently a zero-sum game, and not everyone can be above-average. Seeing school performance in such terms, divorced from classroom practice and public policy, makes both meaningful praise *and* criticism of schools very difficult. Moreover, the constant reinforcement of the myth of declining school performance will continue the erosion of support for the good schools that exist and make intense discussion of the needs of children more difficult.

Return to Table of Contents

Where To Go

Some alternative models of accountability may reverse the destructive tendencies of statistical accountability systems, both in political and practice terms. Reconstructing public education in its best sense (schooling *for* children, their families, and the public) requires

connecting schools in a meaningful and explicitly political way with broader communities. In the same way that the development of the Central Park East elementary and secondary schools under Deborah Meier's leadership required both bureaucratic support and political connections to survive and thrive (Fliegel 1993; Meier 1995), so other schools and school critics dissenting from the current accountability trend must craft an alternative support structure, both within and extending beyond public schooling. Sizer (1992) argues for opening up schools to external evaluation for pedagogical reasons, to keep teachers in touch with reasonable expectations of what students should do. In addition, allowing friendly critics into schools serves an explicitly political purpose, giving community members a concrete sense of what happens in schools. No statistics can substitute for the type of immediate contact such external evaluation provides.

Permitting external evaluation is difficult today. Allowing strangers into schools is threatening because it erodes, at least on a symbolic level, the commitment to professional autonomy which administrators have maintained for almost one hundred years. In practical terms, it requires balancing the legitimate needs of teachers for enough time to plan and try out ideas against the interests of parents and the public to know what is happening in schools. In systems where many teachers may be from ethnic and racial groups different from their students, the tension between teachers and parents may be real, and letting parents into evaluation may be politically tricky (B. Peterson 1997). Yet educators must acknowledge the need to move beyond professionalism as the primary route to support for public schools. Isolating the workings of schools from the public has done teachers and administrators a disservice in the long term as professionalism has declined as a successful route to status and autonomy.

External community evaluation is not the only conceivable way of crafting alternatives to high-stakes standardized test accountability. Others might meet the same needs (e.g., Bernauer and Cress 1997). Common to solving the political problems of accountability are the following three requirements:

- Accountability should encourage deeper discussion of educational problems. Student performance should be the starting point of educational politics, not an occasion for political opportunism or crude comparisons. Statistical accountability, with the centralization of statistical production and dissemination through popular news sources, encourages oversimplification rather than a more extensive public discussion.

- Accountability should connect student performance with classroom practice. Statistical accountability, with the abstraction of student performance into numbers without context, removes classroom practices from the discussion of educational reform.

- Accountability should make the interests of all children common. This sense of commonality is the best meaning of "public" in public schooling. Statistical accountability systems intensify educational triage, encouraging schools to isolate and devote fewer resources to students whom schools judge as difficult to teach. Politically, statistical accountability systems divide the interests of schools and communities through competition for prestige and resources.

No one should pretend that accountability is without conflict or unproblematic. We should face those conflicts and issues directly, however, instead of hiding behind existing standardized testing. Some parents and others may well see statistical comparisons as a primary way for them to gauge school programs and children's education, or as a way to advance specific interests. For example, parents of students with disabilities and disability advocates face real quandaries over accountability. On the one hand, high-stakes testing has created incentives for segregating students (McGill-Franzen and Allington 1993). On the other hand, the national rhetoric emphasizing achievement for *all* students has provided a

lever to criticize the omission of students with disabilities from assessment systems, to craft new federal law encouraging inclusion in assessment, and to create guidelines for state officials seeking to change assessment practices (Thurlow, Elliott, Ysseldyke, and Erickson 1996; also visit the National Center on Educational Outcomes site at http://www.coled.umn.edu/NCEO/). This dilemma is rooted in the tension between wanting to protect students with disabilities from the deleterious consequences of high-stakes testing and yet also wanting whatever accountability systems exist to pay attention to their interests. Those criticizing statistical accountability systems must understand this and similar dilemmas of parents and advocates. Changing attitudes and assumptions, while protecting what many see as important in statistical accountability, requires modeling of worthwhile alternatives and small-scale demonstrations that are explicitly political. Over time, if not immediately, schools need a plausible, fair way to evaluate school improvement. With enough local models of alternative accountability, then perhaps the dynamics of educational politics at state and national levels can change to become broader, connect with classroom practices, and require more than sound bites. Without those concrete examples, however, the domination of crude statistical evaluation of schools will continue, to the detriment of schools, children, their families, and the public.

Return to Table of Contents

Acknowledgements

I am indebted to Gene Glass and anonymous reviewers at *Educational Policy Analysis Archives* , Brian Balogh, Douglas Fuchs, and the students in my masters history of education class for contributions to my thinking about this subject. I am solely responsible, of course, for the interpretations here and the accuracy of detail.

Notes

1. I mean by standardized tests those administered in whole-group settings with quantifiable results. These include multiple-choice tests and also performance-based tests whose results are reportable in quantifiable terms. Thus, Advanced Placement exams conducted by Educational Testing Service are standardized tests for the purposes of this article because, even though parts of the test are performance-based (such as essays), the essays are scored by a quantifiable rubric system and the whole test reported on the company's 1-5 scale for such tests. Moreover, reporting scores by numbers allows the simplified public discussion which is my focus here. For an introduction to Lauren Resnick's advocacy of measurement-driven reform, see Simmons and Resnick (1993). For issues involved in Kentucky and Arizona (respectively), see Jones and Whitford (1997) and Noble and Smith (1994).

2. An anonymous reviewer noted that the line between practice and political legacies is fuzzy. In many ways, the debates over census undercounting and the consumer price index are also debates about the political rhetoric of the reapportionment process and future support for government entitlement programs. Nonetheless, the distinction between the two legacies is a useful heuristic device for explaining why the literature on perverse incentives of high-stakes testing does not address the critical issue of school politics.

3. According to the Vanderbilt Television News Archives, the following broadcasts discussed standardized test score levels between 1968 and 1987: October 28, 1975 (ABC, CBS); November 17, 1975 (CBS); August 23, 1977 (ABC, CBS); August 24, 1977 (ABC, commentary); September 1, 1977 (CBS, commentary); September 21, 1982 (CBS); September 19, 1984 (ABC); January 9, 1985 (NBC); January 26, 1985 (CBS, NBC); September 22, 1987 (NBC). The search terms included "standardized

and test*," "test and scor*," "SAT and scor*," and "SAT and (college or scholastic)." Excluded from this list are stories about the alleged discriminatory nature of tests.

4. I agree with Stedman (1996a, 1996b) that schools are not as good as they should be. Those problems do not mean the myth of declining quality is true: schools have been inconsistent and too often mediocre for many years.

References

Note: Links here use a second browser window.

Abbott v. Burke . 1997. 149 N.J. 145. Retrieved September 27, 1997 from the World Wide Web: http://www-camlaw.rutgers.edu/decisions/supreme/m-622-96.opn.html

Agostini v. Felton . 1997. 117 U.S. 1997. Retrieved September 27, 1997 from the World Wide Web: http://supct.law.cornell.edu/supct/html/96-552.ZS.html

Axinn, June, and Mark J. Stern. 1988. *Dependency and poverty: Old problems in a new world* . Lexington, Mass: Lexington Books.

Balogh, Brian. 1991a. *Chain reaction: Expert debate and public participation in American commercial nuclear power, 1945-1975* . New York: Cambridge University Press.

Balogh, Brian. 1991b. Reorganizing the organizational synthesis: Reconsidering modern American federal-professional relations. *Studies in American Political Development* 5 (1): 119-172.

Berliner, David C., and Bruce J. Biddle. 1995. *The manufactured crisis: Myths, fraud, and the attack on America's public schools* . Reading, Mass.: Addison-Wesley.

Berliner, David C., and Bruce J. Biddle. 1996. Making molehills out of molehills: Reply to Lawrence Stedman's review of *The manufactured crisis* . *Educational Policy Analysis Archives* 4 (February 26). Retrieved September 27, 1997, from the World Wide Web: http://olam.ed.asu.edu/epaa/v4n3.html

Bernauer, James A., and Katherine Cress. 1997. How school communities can help redefine accountability assessment. *Phi Delta Kappan* 79 (September): 71-75.

Bock, R. Darrell, and Richard Wolfe. 1996. *Audit and review of the Tennessee Value-Added Assessment System (TVAAS): Final report* . Nashville, Tenn.: Tennessee Comptroller of the Treasury.

Bodenheimer, Thomas. 1996. HMO backlash -- righteous or reactionary? *New England Journal of Medicine* 335: 1601-1604.

Bracey, Gerald W. 1991. Why can't they be like we were? *Phi Delta Kappan* 73: 104-17.

Bracey, Gerald W. 1992. The second Bracey report on the condition of public education. *Phi Delta Kappan* 74: 104-8, 110-17.

Bracey, Gerald W. 1993. The third Bracey report on the condition of public education. *Phi Delta Kappan* 75: 104-12, 114-18.

Bracey, Gerald W. 1994. The fourth Bracey report on the condition of public education. *Phi Delta Kappan* 76: 115-27.

Bracey, Gerald W. 1995a. The fifth Bracey report on the condition of public education. *Phi Delta Kappan* 77: 149-60.

Bracey, Gerald W. 1995b. *Final exam: A study of the perpetual scrutiny of American education* .

Bracey, Gerald W. 1996. The sixth Bracey report on the condition of public education. *Phi Delta Kappan* 78: 127-38.

Bracey, Gerald W. 1997. The seventh Bracey report on the condition of public education. *Phi Delta Kappan* 79: 120-36.

Calhoun, Daniel Hovey. 1973. *The intelligence of a people* . Princeton, N.J.: Princeton University Press.

Cannell, John Jacob. 1989. *How public educators cheat on standardized achievement tests: The "Lake Wobegon" report* . ERIC Reproduction Document No. ED 314 454.

Cohen, Sol, ed. 1973. *Education in the United States: A documentary history* . New York: Random House.

Darling-Hammond, Linda. 1988. The futures of teaching. *Educational Leadership* 46 (November): 4-10.

Darling-Hammond, Linda. 1992. Educational indicators and enlightened policy. *Educational Policy* 6 (September): 235-65.

Darling-Hammond, Linda, and Carol Ascher. 1991. *Creating accountability in the big city school systems* . Urban Diversity Series No. 102. New York: National Center for Restructuring Education, Schools, and Teaching. ERIC Reproduction Document No. ED 334 339.

Darling-Hammond, Linda, Arthur E. Wise, and Stephen P. Klein. 1995. *A license to teach: Building a profession for 21st-century schools* . Boulder, Colo.: Westview Press.

Delpit, Lisa. 1995. *Other people's children: Cultural conflict in the classroom* . New York: The New Press.

Diegmueller, Karen, and Millicent Lawton. 1996. The road not taken. *Education Week* (April 24). Retrieved September 27, 1997, from the World Wide Web: http://www.edweek.com

Douglas, Davison M. 1995. *Reading, writing, and race: The desegregation of the Charlotte schools* . Chapel Hill, N.C.: University of North Carolina Press.

Education Commission of the States. 1997. *Accountability: State Policies* . Denver: Education Commission of the States. Retrieved September 27, 1997 from the World Wide Web: http://www.ecs.org/ecs/24aa.htm

Educational Improvement Act. 1992. Tennessee Acts Chapter 353.

Educational Improvement Acts Amendments. 1997. Tenn. Acts Chapter 434. Retrieved September 27, 1997 from the World Wide Web: http://www.legislature.state.tn.us/bills/100gahtm/100_chap/pubc0434.htm

Fine, Michelle. 1991. *Framing dropouts: Notes on the politics of an urban public high school* . Albany, N.Y.: State University of New York Press.

Finn, Michael. 1997. State school systems bill passes Senate. *Chattanooga Free Press* (May 31).

Fisher, Thomas H. 1996. *A review and analysis of the Tennessee Value-Added Assessment System: Part II* . Nashville, Tenn.: Comptroller of the Treasury.

Fliegel, Sy. 1993. *Miracle in East Harlem: The fight for choice in public education* . New York: Times Books.

Friedson, Eliot. 1984. Are professions necessary? In *The authority of experts* , edited by Thomas L. Haskell (pp. 3-27). Bloomington, Ind.: Indiana University Press.

Fuchs, Douglas, and Lynn S. Fuchs. 1995. Special education can work. In *Issues in educational placement: Students with emotional and behavioral disorders* (pp. 363-77), edited by James M. Kauffman, John W. Lloyd, Daniel P. Hallahan, and Terry A. Astuto. Hillsdale, N.J.: Erlbaum.

Furstenberg, Frank. 1991. As the pendulum swings: Teenage childbearing and social concern. *Family Relations* 40: 127-38.

Galambos, Louis. 1970. The emerging organizational synthesis in modern American history. *Business History Review* 44: 279-290.

Ginsburg, Alan L, Jay Noell, and Valene White Plisko. 1988. Lessons from the wall chart. *Educational Evaluation and Policy Analysis* 10 (Spring): 1-12.

Glass, Gene V 1990. Using student test scores to evaluate teachers. In *The new handbook of teacher evaluation* (pp. 229-40), edited by Jason Millman and Linda Darling-Hammond. Newbury Park, Calif.: Sage Publications.

Haefele, Donald L. 1992. Evaluating teachers: An alternative model. *Journal of Personnel Evaluation in Education* 5: 335-45.

Hess, G. Alfred. 1991. *School restructuring, Chicago style* . Newbury Park, Calif.: Corwin Press.

Hillsborough County Public Schools. 1997. *Our students' test scores reflect academic achievement* . Tampa, Fla.: Office of Communications and Governmental Relations.

Hoff, David J. 1997. White House, GOP craft agreement on testing. *Education Week* (November 12): 1, 23.

Holmes Group. 1986. *Tomorrow's teachers* . East Lansing, Mich.: Holmes Group.

How midstate schools stack up. 1994a. Nashville *Tennessean* (October 15): 4A.

How midstate schools stack up. 1994b. Nashville *Tennessean* (October 16): 6A-7A.

Johnston, Robert C. 1997. Just saying no. *Education Week* (April 9). Retrieved September 27, 1997, from the World Wide Web: http://www.edweek.com

Jones, Ken, and Betty Lou Whitford. 1997. Kentucky's conflicting reform principles: High-stakes school accountability and student performance assessment. *Phi Delta Kappan* 79: 276-81.

Jones, Lyle V. 1996. A history of the National Assessment of Educational Progress and some questions about its future. *Educational Researcher* 25 (October): 15-22.

Justiz, Manuel J., and Marilyn C. Kameen. 1988. Increasing the representation of minorities in the teaching profession. *Peabody Journal of Education* 66 (Fall): 91-100.

Kaestle, Carl F. 1973. *Joseph Lancaster and the monitorial school system: A documentary history* . New York: Teachers College Press.

Katz, Michael B. 1992. Chicago school reform as history. *Teachers College Record* 94: 56-72.

Katznelson, Ira, and Margaret Weir. 1985. *Schooling for all: Class, race, and the decline of the democratic ideal* . New York: Basic Books.

Kirst, Michael W., and Christopher Mazzeo. 1996. The rise, fall, and rise of state assessment in California, 1993-1996. Paper presented at the Annual Meeting of the American Educational Research Association. ERIC Reproduction Document Number ED 397 133.

Klausnitzer, Dorris. 1996. Pupils need practice on that first "R." Nashville *Tennessean* (November 15): 1A, 8A-10A. The *Tennessean* 's statistical analysis retrieved September 27, 1997 from the World Wide Web: http://www.tennessean.com/schools/ (defunct link)

Koretz, Daniel. 1992a. NAEP and national testing: Issues and implications for educators. *NASSP Bulletin* 76 (September): 30-40.

Koretz, Daniel. 1992b. What happened to test scores, and why? *Educational Measurement: Issues and Practice* 11 (Winter): 7-11.

Koretz, Daniel, and Edward Diebert. 1993. *Interpretations of National Assessment of Educational Progress (NAEP) anchor points and achievement levels by the print media in 1991* . Santa Monica, Calif.: The RAND Corporation. ERIC Document Reproduction No. ED 367 683.

Kozol, Jonathan. 1991. *Savage inequalities: Children in America's schools* . New York: Crown Publishers.

Kyle, Charles, and Edward Kantowicz. 1991. Bogus statistics. *Latino Studies Journal* 2 (May): 34-52.

Labaree, David F. 1988. *The making of an American high school: The credentials market*

and the Central High School of Philadelphia, 1838-1939 . New Haven, Conn.: Yale University Press.

Labaree, David F. 1992. Power, knowledge, and the rationalization of teaching: A genealogy of the movement to professionalize teaching. *Harvard Educational Review* 62: 123-54.

Lawton, Millicent. 1997. Riley delays national tests' development. Education Week (October 1). Retrieved October 2, 1997, from the World Wide Web: http://www.edweek.com

Lipsky, Michael. 1980. *Street-level bureaucracy: Dilemmas of the individual in public service* . New York: Russell Sage Foundation.

Madaus, George F. 1988. The distortion of teaching and testing: high-stakes testing and instruction. *Peabody Journal of Education* 65 (Spring): 29-46.

Madaus, George F. 1991. The effects of important tests on students. *Phi Delta Kappan* 73: 226-31.

Maeroff, Gene. 1976. Aptitude test lag is puzzling experts. *New York Times* (September 12): 17.

McDonnell, Lorraine M. 1997. *The politics of state testing: Implementing new student assessments* . National Center for Research on Evaluation, Standards, and Student Testing (CRESST) Technical Report 424. Los Angeles, Calif.: CRESST. Retrieved November 14, 1997, from the World Wide Web: http://cresst96.cse.ucla.edu/Reports/tech424.pdf

McGill-Franzen, Anne, and Richard L. Allington. 1993. Flunk'em or get them classified: The contamination of primary grade accountability data. *Educational Researcher* 22 (January-February): 19-22.

McGrew, Kevin S., Mike L. Vanderwood, Martha L. Thurlow, and James E. Ysseldyke. 1995. Why we can't say much about the status of students with disabilities during education reform. NCEO Synthesis Report 21. Minneapolis, Minn.: National Center on Educational Outcomes. ERIC Reproduction Document Number ED 396 475.

Meers, Walter R. 1997. Infighting breaks out over census. CNN News (October 2). Retrieved October 2, 1997, from the World Wide Web: http://www.allpolitics.com/1997/10/02/ap/census (defunct link)

Meier, Deborah. 1995. *The power of their ideas: Lessons for America from a small school in Harlem* . Boston: Beacon Press.

Meier, Deborah. 1997. How our schools could be: Standards, top-down mandates, and grass-roots communities. *Rethinking Schools* 11 (Summer): 8-9.

Merrow, John. 1997. *Testing . . . testing . . . testing* . Public Broadcasting System. Script retrieved September 27, 1997 from the World Wide Web: http://www.pbs.org/merrow/tttscript.txt

Miller, Matthew. 1997. Surprise! National school standards exist. *U. S. News and World Report* (November 17). Retrieved December 23, 1997, from the World Wide Web: http://www.usnews.com/usnews/issue/971117/17stan.htm

Milliken v. Bradley . 1974. 418 U.S. 717.

Missouri v. Jenkins . 1995. 515 U.S. 70. Retrieved September 27, 1997, from the World Wide Web: http://supct.law.cornell.edu/supct/html/93-1823.ZS.html

National Center on Educational Outcomes. 1994. *Educational accountability for students with disabilities* . NCEO Policy Directions Number 3. Minneapolis, Minn.: National Center on Educational Outcomes. ERIC Reproduction Document No. ED 378 775. Retrieved November 19, 1997, from the World Wide Web: http://www.coled.umn.edu/NCEO/OnlinePubs/Policy3.html

National Commission on Excellence in Education. 1983. *A nation at risk* . Washington, D.C.: Government Printing Office. Retrieved September 27, 1997, from the World Wide Web: http://www.ed.gov/pubs/NatAtRisk/index.html

Neill, Monty. 1996. Assessment reform at crossroads. *Education Week* (February 21). Retrieved September 27, 1997, from the World Wide Web: http://www.edweek.com

Noble, Audrey J., and Mary Lee Smith. 1994. Old and new beliefs about measurement-driven reform: "Build it and they will come." *Educational Policy* 8 (June): 111-36.

Ohio Association of Independent Schools v. Goff . 1996. 92 F.3d 419 (6th Cir.). Retrieved September 27, 1997 from the World Wide Web: http://www.ljextra.com/cgi-bin/f_cat?prod/ljextra/data/external/1996/08/96080l2.c06

Orfield, Gary. 1969. *The reconstruction of southern education: The schools and the 1964 Civil Rights Act* . New York: John Wiley & Sons.

Orfield, Gary. 1993. *The growth of segregation in American schools: Changing patterns of separation and poverty since 1968* . Alexandria, Va.: National School Boards Assocation, Council of Urban Boards of Education.

Orfield, Gary, Susan E. Eaton, and the Harvard Project on School Desegregation. 1996. *Dismantling desegregation: The quiet reversal of Brown v. Board of Education* . New York: New Press.

Peterson, Bob. 1997. We need a new vision of teacher unionism. *Rethinking Schools* 11 (Summer): 1, 3-5.

Peterson, Kenneth D. 1995. *Teacher evaluation: A comprehensive guide to new directions and practices* . Thousand Oaks, Calif.: Corwin Press, Inc.

Ravitch, Diane. 1983. *The troubled crusade: American education, 1945-1980* . New York: Basic Books.

Ravitch, Diane. 1995. *National standards in American education: A citizen's guide* . Washington, D.C.: Brookings.

Reese, William J. 1995. *The origins of the American high school* . New Haven, Conn.: Yale University Press.

Rose, Lowell C., Alec M. Gallup, and Stanley M. Elam. 1997. The 29th annual Phi Delta

Kappa/Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan* 79 (September): 41-56.

Roush, Wade. 1996. A census in which all Americans count. *Science* 274: 713-14.

Ruggles, Patricia. 1990. *Drawing the line: Alternative poverty measures and their implications for public policy* . Washington, D.C.: Urban Institute Press.

Sanders model to measure "value added." 1991. *TEA News* 22 (May): 5.

Sanders, William L., and Sandra P. Horn. 1994. The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education* 8: 299-311.

Sapon-Shevin, Mara. 1993. Gifted education and the protection of privilege: Breaking the silence, opening the discourse. In *Beyond silenced voices: Class, race, and gender in United States Schools* (pp. 25-44), edited by Lois Weis and Michelle Fine. Albany, N.Y.: State University of New York Press.

Schell, Jonathan. 1975. *The time of illusion* . New York: Knopf.

Shepard, Lorrie A. 1991. Will national tests improve student learning? *Phi Delta Kappan* 73: 232-38.

Simmons, Warren, and Lauren Resnick. 1993. Assessment as the catalyst of school reform. *Educational Leadership* 50 (February) 11-15.

Sizer, Theodore R. 1992. *Horace's school: Redesigning the American high school* . Boston: Houghton Mifflin Co.

Sizer, Theodore R. 1995. Will national standards and assessments make a difference? In *Debating the future of American education: Do we need national standards and assessments?* (pp. 33-39), edited by Diane Ravitch. Washington, D.C.: Brookings Institution.

Skocpol, Theda. 1991. Targeting within universalism: Politically viable policies to combat poverty in the United States. In *The urban underclass* (pp. 411-436), edited by Christopher Jencks and Paul E. Peterson. Washington, D.C.: Brookings Institution.

Smith, Mary Lee. 1991. Put to the test: The effects of external testing on teachers. *Educational Researcher* 20 (June-July): 8-11.

Smith, Mary Lee, and Claire Rottenberg. 1991. Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice* 10 (Winter): 7-11.

Spring, Joel. 1989. *The sorting machine revisited: National educational policy since 1945* (updated ed.). New York: Longman.

Starr, Paul. 1982. *The social transformation of American medicine* . New York: Basic Books.

Starr, Paul. 1987. The sociology of official statistics. In *The politics of numbers* (pp. 7-57), edited by Paul Starr and William Alonso. New York: Russell Sage Foundation.

Stedman, Larence A. 1996b. Respecting the evidence: The achievement crisis remains real. *Educational Policy Analysis Archives* 4 (April 4). Retrieved September 27, 1997 from the World Wide Web: http://olam.ed.asu.edu/epaa/v4n7.html

Stedman, Lawrence A. 1996a. The achievement crisis is real: A review of *The manufactured crisis* . *Educational Policy Analysis Archives* 4 (January 23). Retrieved September 27, 1997 from the World Wide Web: http://olam.ed.asu.edu/epaa/v4n1.html

Strike, Kenneth A. 1990. The ethics of educational evaluation. In *The new handbook of teacher evaluation* (pp. 356-73), edited by Jason Millman and Linda Darling-Hammond. Newbury Park, Calif.: Sage Publications.

Tennessee Department of Education. 1996. 21st Century Report Card. Retrieved September 27, 1997 from the World Wide Web: http://www.state.tn.us/education/rptcrd96/index.html

Thurlow, Martha L., Judy Elliott, James E. Ysseldyke, and Ron Erickson. 1996. *Questions and answers: Tough questions about accountability systems and students with disabilities* . Minneapolis, Minn.: National Center on Educational Outcomes. ERIC Reproduction Document No. ED 404 802.

Tyack, David B. 1974. *The one best system: A history of American urban education* . Cambridge, Mass.: Harvard University Press.

Tyack, David B., and Elisabeth Hansot. 1982. *Managers of virtue: Public school leadership in America, 1820-1980* . New York: Basic Books.

Tyack, David B., and Larry Cuban. 1995. *Tinkering toward utopia: A century of public school reform* . Cambridge, Mass.: Harvard University Press.

Vinvoskis, Maris. 1988. *An "epidemic" of adolescent pregnancy? Some historical and policy considerations* . New York: Oxford University Press.

Weick, Karl E. 1976. Educational organizations as loosely coupled systems. *Administrative Science Quarterly* 21: 1-19.

Weiss, Carol H. 1988. Interview study. In *Reporting of social science in the national media* (pp. 21-171), edited by Carol H. Weiss and Eleanor Singer. New York: Russell Sage Foundation.

Weiss, Janet A., and Judith E. Gruber. 1987. The managed irrelevance of federal education statistics. In *The politics of numbers* (pp. 363-91), edited by Paul Starr and William Alonso. New York: Russell Sage Foundation.

Werum, Regina. 1997. Sectionalism and racial politics: Federal vocational policies and programs in the predesegregation South. *Social Science History* 21: 399-453.

Wiebe, Robert H. 1967. *The search for order, 1877-1920* . New York: Hill and Wang.

Will Washington cut our COLA? 1997. *Solidarity* (January-February). Retrieved October 1, 1997, from the World Wide Web: http://www.uaw.org/solidarity/9701/03b.html

Wilson, Thomas A. 1996. *Reaching for a better standard: English school inspection and the*

dilemma of accountability for American public schools . New York: Teachers College Press.

Wirth, Arthur G. 1992. *Education and work for the year 2000: choices we face* . San Francisco: Jossey-Bass.

Wise, Arthur E. 1979. *Legislated learning: The bureaucratization of the American classroom* . Berkeley, Calif.: University of California Press.

Wise, Arthur E. 1994. Choosing between professionalism and amateurism. *Educational Forum* 58: 139-46.

Wise, Arthur E., and Jane Leibbrand. 1993. Accreditation and the creation of a profession of teaching. *Phi Delta Kappan* 75: 133-36, 154-57.

Return to Table of Contents

About the Author

Sherman Dorn

dorn@typhoon.coedu.usf.edu

Sherman Dorn is Assistant Professor in the Department of Psychological and Social Foundations at the University of South Florida. He received his Ph.D. in history at the University of Pennsylvania in 1992 based on his work on the history of dropout policies. He is currently looking at the history of special education in Nashville, Tennessee, from 1940 to 1990.

Copyright 1998 by the *Education Policy Analysis Archives*

The World Wide Web address for the *Education Policy Analysis Archives* is http://olam.ed.asu.edu/epaa

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-2411. (602-965-2692). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu . The Commentary Editor is Casey D. Cobb: casey@olam.ed.asu.edu .

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Arlen Gullickson
Western Michigan University

Aimee Howley
Marshall University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Thomas Mauhs-Pugh
Rocky Mountain College

William McInerney
Purdue University

Les McLean
University of Toronto

Anne L. Pemberton
apembert@pen.k12.va.us

Richard C. Richardson
Arizona State University

Dennis Sayers
University of California at Davis

Michael Scriven
scriven@aol.com

Robert Stonehill
U.S. Department of Education

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina--Greensboro

Benjamin Levin
University of Manitoba

Dewayne Matthews
Western Interstate Commission for Higher Education

Mary P. McKeown
Arizona Board of Regents

Susan Bobbitt Nolen
University of Washington

Hugh G. Petrie
SUNY Buffalo

Anthony G. Rud Jr.
Purdue University

Jay D. Scribner
University of Texas at Austin

Robert E. Stake
University of Illinois--UC

Robert T. Stout
Arizona State University